IJSPT

ORIGINAL RESEARCH

INTRA- AND INTER-RATER RELIABILITY OF THE SELECTIVE FUNCTIONAL MOVEMENT ASSESSMENT (SFMA) IN HEALTHY PARTICIPANTS

Justin M. Stanek, Ed.D., ATC[1]
Joshua Smith, PT, DPT[2]
Jake Petrie, SPT, ATC[1]

ABSTRACT

Background: The Selective Functional Movement Assessment (SFMA) is a popular assessment tool used to observe and detect components of dysfunctional movement patterns. The goal of the assessment is to identify impairments throughout the kinetic chain that may be contributing to movement dysfunction and/or pain.

Hypothesis/Purpose: The purpose of this research was to determine the intra- and inter-rater reliability of the 10 top-tier movements of the SFMA using the categorical scoring system. It was hypothesized the intra- and inter-rater reliability of the SFMA would be acceptable with variations based on the objectivity of the scoring criteria and the experience of the rater.

***Study Design*:** Cross-sectional reliability study.

Methods: 25 (17 male, 8 female), physically active participants (age: 21.2 ± 1.6years, height: 177.1 ± 10.7cm, weight: 74.9 ± 13.9kg) were independently assessed in real time by three clinicians during two separate visits to the lab using a standard instructional script. Clinicians had varying levels of experience with the SFMA and the two visits occurred a minimum of 48 hours and maximum of seven days apart. Results from each clinician were compared within and between raters using the Kappa coefficient and ratings of absolute agreement.

Results: Overall, slight to substantial intra- and inter-rater reliability were observed using the categorical scoring tool, although variations existed depending on the movement pattern. Kappa coefficients for intra-rater reliability ranged from 0.21-1.00, while % absolute agreement ranged from 0.64-1.00. Inter-rater reliability for the same measures ranged from 0.11-0.89 and 0.52-0.96 respectively. Clinicians certified in the SFMA with the greatest amount of experience using the SFMA demonstrated higher intra-rater reliability. Similarly, higher inter-rater reliability was found between certified raters with the most experience.

Conclusions: Certified SFMA raters with greater amounts of experience can demonstrate adequate intra- and inter-rater reliability using the categorical scoring method.

Level of Evidence: 2, Reliability study

Keywords: Dysfunction, functional movement, movement screen, movement system, repeatability

[1] Illinois State University, Normal, IL, USA
[2] IVYREHAB Physical Therapy, Bloomington, IL, USA

The authors do not have any conflicts of interest to disclose regarding the research.

CORRESPONDING AUTHOR
Justin M. Stanek
Illinois State University
Campus Box 5120
Normal, IL 61790-5120
(309) 438-5862—phone
(309) 438-5559—fax
E-mail: jmstane@ilstu.edu

INTRODUCTION

The Selective Functional Movement Assessment (SFMA) is one of the many tools used by health care professionals to observe movement restrictions in individuals with known musculoskeletal injuries. While the reliability of the Functional Movement Screen™ (FMS™) has been heavily studied, only two published studies have examined the SFMA.[1,2] The SFMA consists of ten functional movement patterns involving both the upper and lower extremities and is scored by a trained clinician based on the quality of movement.[3] Using this assessment, the clinician can identify dysfunctional and/or painful movement patterns and detect components of the patterns to determine the possible causes. This differentiates the SFMA as an assessment tool used for evaluating injured subjects as opposed to the FMS™ which is designed to screen healthy participants. Ultimately, when using the movement assessment, the clinician is able to identify if the dysfunctional movement is caused by tissue extensibility, joint mobility, or stability/motor control dysfunction.[3,4] At the core of the assessment is the concept of regional interdependence.[5] This notion allows the clinician to identify impairments within the kinetic chain that may seem unrelated to the primary complaint, but may contribute to their movement dysfunction.[3]

Similar to the FMS™, the SFMA must be determined to be reliable prior to investigating other factors such as corrective exercises or specific treatment interventions. Multiple investigations have examined the reliability of the FMS™ under a variety of contexts and two recent systematic reviews found the screen to be a reliable method for evaluating motion.[6,7] However, unlike the FMS™, the SFMA is designed to be used in the presence of pain as part of a comprehensive musculoskeletal evaluation. It is paramount that clinicians can trust that the results of their SFMA exam are reliable when making decisions regarding patient status, progression, and discharge.

Previous SFMA reliability studies examined healthy participants using video analysis[1] or unhealthy participants using both video analysis and real-time scoring.[2] Prior to implementing the assessment as part of a clinician's practice, it is necessary to establish the real-time reliability of the assessment on a healthy population since the individual's pain and/or dysfunction could alter movement patterns between sessions and affect scoring. Furthermore, the SFMA was designed to be scored real-time and many clinicians do not have the time to review videos of their patients. Therefore, the primary purpose of this study was to examine the inter- and intra-rater reliability of the 10 top-tier SFMA tests using real-time scoring among raters of different education and experience levels with the assessment in a healthy population. The authors hypothesized that movements with objective benchmarks would be scored more reliably than movements that contained subjective criteria. Furthermore, the authors hypothesized that raters with more experience would exhibit higher intra-rater reliability scores.

METHODS

Participants

A convenience sample of 25 (17 male, 8 female) physically active, college-aged participants volunteered to participate in the study. The average age, height, and mass of the participants were 21.2 ± 1.6years, 177.1 ± 10.7cm, and 74.9 ± 13.9kg respectively. All participants were asked to complete a basic pre-participation questionnaire, which included age, sex, height, weight, previous musculoskeletal injuries, and activity level per week. To be included, participants needed to be physically active for ≥30min on three or more days per week. Participants were excluded if they were currently suffering from any musculoskeletal injury that affected their physical activity participation, had undergone any surgery within the prior six months, self-reported any neurologic conditions, or were pregnant. Before the second testing session, examiners reviewed participant's health history information to ensure no changes had occurred. A healthy population was recruited for the study in an attempt to establish the reliability of the assessment without the potential influence of pain and/or dysfunction. Prior to beginning data collection, the study was reviewed and approved by the university's institutional review board. All participants provided written informed consent prior to study participation.

Instrumentation

Participants were scored in real time using the 10 top-tier patterns of the SFMA during two visits to the lab, each separated by a minimum of 48 hours and maximum of seven days. The SFMA is comprised of the following fundamental movement patterns: 1) Cervical flexion, 2) Cervical extension, 3) Cervical rotation, 4) Upper extremity pattern 1 (medial rotation, adduction, extension), 5) Upper extremity pattern 2 (lateral rotation, abduction, flexion) 6) Multi-segmental flexion, 7) Multi-segmental extension, 8) Multi-segmental rotation, 9) Single leg balance, 10) Overhead deep squat. Each movement was scored categorically based on function and pain into one of four categories (Functional non-painful-FN, Functional painful-FP, Dysfunctional non-painful-DN, or Dysfunctional painful-DP). Participants were not familiar with or taught the grading criteria for any of the movements. A standard verbal script was read to each participant for the desired movement by the same rater (Appendix A). Participants were instructed to alert the raters of any pain experienced during the 10 movements. The movement instructions and scoring criteria were based on those presented in the SFMA Level 1 course[4] and detailed in previous publications.[1,3]

Raters included one athletic training faculty member with 15 hours of SFMA training, SFMA Level 1 certification, and two years' experience scoring the SFMA (Rater A), one physical therapist with 15 hours of SFMA training, SFMA Level 1 certification, and one year of experience with SFMA (Rater B), and one athletic training student without SFMA certification and no formal training (Rater C). The undergraduate athletic training student had no formal training with the SFMA, however, he did complete a summer internship and worked closely with a physical therapist that used the SFMA on a daily basis. All raters were provided up-to-date, detailed instructions on the SFMA scoring prior to the study. All three raters were present during the data collection sessions and were allowed to move freely about to evaluate each participant's movement. If needed, raters were allowed to ask the participant to repeat the movement pattern and all raters were allowed an opportunity to re-score the movement. Raters did not confer prior to evaluating the movement and each rater was blinded to the scoring of the other raters. Furthermore, the raters used new scoring sheets for each participant, therefore, the raters were unaware of how they scored the movement during the participant's first visit.

Procedures

Participants arrived to the lab and the informed consent was obtained from each participant prior to testing. Following completion of the informed consent, participants completed a pre-participation questionnaire that included demographic information and evaluated the participant for the inclusion/exclusion criteria. The standard instructional script was read to each participant for all movements during both visits. The primary investigator (Rater A) also provided a visual demonstration of each movement prior to the participants' attempt. If necessary, the script was repeated and each of the raters visually verified the movements for correct execution. Movements were scored in the order presented by the SFMA manual,[4] beginning with the cervical flexion test and ending with the overhead deep squat. All three raters performed their ratings upon completion of the subject. Evaluators did not cue the participant during any of the movements, nor did they discuss how they scored the movements at any time during or after data collection.

Statistical Analyses

All raw data were initially entered into Microsoft Excel 2016 (Microsoft, Redmond, VA). Data were reduced and copied to SPSS (version 21; IBM Corporation, Armonk, NY) for statistical analysis. Because the goal of the study was to establish the reliability of the individual movement pattern, the authors were not concerned about the movement direction or limb (right/left). Therefore, movements that had a right/left component were pooled together for analysis (i.e. left and right cervical rotation pooled together as cervical rotation). Categorical scores were compared within and between raters (A-B, A-C, and B-C) using both absolute agreement and the Kappa coefficient. The 95% confidence interval (CI) for the Kappa coefficient was calculated using the formula recommended by McHugh[8] ($k - 1.96 \times SE_k$ to $k + 1.96 \times SE_k$). The strength of agreement was assessed using the Kappa coefficient and the interpretation has been previously described as ≤0.1 = poor, 0.1-0.2 = slight, 0.21-0.4 = fair, 0.41-0.6 = moderate, and 0.61-0.8 = substantial, and 0.81-1.0 = almost perfect.[9]

RESULTS

All 25 participants completed both testing sessions and their data were used for analysis. None of the participants experienced pain during any of the movements. As a group, participants were scored FN most frequently on the cervical flexion pattern (79%) followed by the upper extremity pattern 2 (lateral rotation and flexion; 77%). Participants scored DN most frequently on the single leg stance (77%) followed by the overhead deep squat (75%) and upper extremity pattern 1 (medial rotation and extension; 73%). Rater B scored the highest number of dysfunctional movements followed by Rater A and C respectively.

Results for intra-rater reliability (Kappa, 95% CI, and % agreement) for each rater are presented in Table 1. Kappa values for all raters ranged from slight to substantial depending on the movement pattern evaluated. Rater A demonstrated the highest intra-rater reliability followed by B, and C respectively. The highest intra-rater reliability for all raters occurred during the overhead deep squat followed by the cervical flexion test. The lowest intra-rater reliability for all raters occurred during the cervical extension test (Figure 1), followed by single-leg stance (Figure 2) and multisegmental extension test (Figure 3). Results for the inter-rater reliability (Kappa, 95% CI, and % agreement) are presented in Table 2. Kappa values for all raters ranged from slight to substantial depending on the movement pattern evaluated. Highest inter-rater reliability occurred between raters A & B, followed by A & C, and B & C respectively. The cervical flexion test showed the highest inter-rater reliability while multisegmental extension (Figure 3) showed the lowest inter-rater reliability, followed by cervical rotation (Figure 4).

Figure 1. *Cervical extension test.*

Table 1. *Intra-rater reliability of the categorical scoring of the SFMA presented as Kappa coefficient (95% CI's) and agreement (%).*

	Rater A		Rater B		Rater C		Mean of All Raters	
	Kappa	Agree	Kappa	Agree	Kappa	Agree	Kappa	Agree
C Flex	0.64 (0.34-0.94)	0.84	0.57 (0.31-0.83)	0.84	0.72 (0.42-1.02)	0.88	**0.64**	**0.85**
C Ext	0.38 (0.03-0.72)	0.68	0.26 (-0.05-0.57)	0.64	0.25 (-0.20-0.70)	0.76	**0.3**	**0.71**
C Rot	0.41 (0.07-0.74)	0.84	0.49 (0.25-0.73)	0.76	0.57 (0.36-0.78)	0.78	**0.49**	**0.79**
MRE	0.61 (0.30-0.92)	0.90	0.57 (0.32-0.82)	0.84	0.46 (0.16-0.76)	0.82	**0.55**	**0.85**
LRF	0.41 (0.01-0.80)	0.90	0.51 (0.25-0.77)	0.82	0.38 (0.06-0.69)	0.80	**0.43**	**0.84**
MSF	0.72 (0.42-1.01)	0.88	0.59 (0.27-0.91)	0.80	0.56 (0.23-0.89)	0.80	**0.62**	**0.83**
MSE	0.53 (0.00-0.76)	0.80	0.21 (-0.05-0.47)	0.68	0.36 (0.00-0.73)	0.68	**0.36**	**0.72**
MSR	0.52 (0.28-0.76)	0.76	0.40 (0.14-0.66)	0.74	0.39 (0.14-0.64)	0.70	**0.4**	**0.71**
SLS	0.40 (0.02-0.80)	0.68	0.39 (0.03-0.75)	0.72	0.45 (0.08-0.82)	0.76	**0.35**	**0.72**
ODS	0.90 (0.50-1.06)	0.96	1.00 (n/a)	1.00	0.63 (0.30-0.96)	0.84	**0.84**	**0.93**
Mean	**0.55**	**0.82**	**0.50**	**0.78**	**0.48**	**0.78**	**0.50**	**0.80**
SD	**0.17**	**0.09**	**0.22**	**0.10**	**0.14**	**0.06**	**0.17**	**0.08**

CI= confidence interval; C Flex= cervical flexion; C Ext= cervical extension; C Rot= cervical rotation; MRE= medial rotation extension; LRF= lateral rotation flexion; MSF= multisegmental flexion; MSE= multisegmental extension; MSR= multisegmental rotation; SLS= single leg stance; ODS= overhead deep squat; SD= standard deviation

Figure 2. *Single leg stance test.*

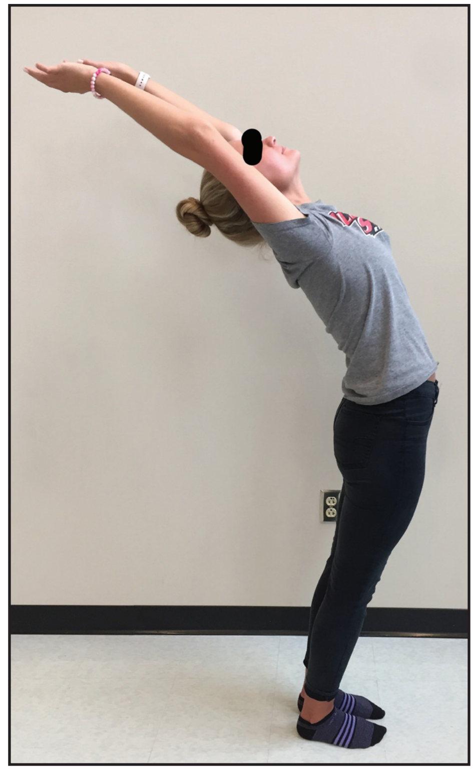

Figure 3. *Multisegmental extension test.*

Table 2. *Inter-rater reliability of the categorical scoring of the SFMA presented as Kappa coefficient (95% CI's) and agreement (%).*

	Rater A-B		Rater A-C		Rater B-C		Mean of All Raters	
	Kappa	Agree	Kappa	Agree	Kappa	Agree	Kappa	Agree
C Flex	0.89 (0.69-1.09)	0.96	0.89 (0.69-1.09)	0.96	0.80 (0.54-1.06)	0.92	**0.86**	**0.95**
C Ext	0.69 (0.42-0.96)	0.84	0.36 (0.01-0.71)	0.72	0.18 (-0.07-0.43)	0.56	**0.41**	**0.71**
C Rot	0.44 (0.22-0.67)	0.76	0.24 (0.08-0.40)	0.58	0.19 (-0.06-0.44)	0.58	**0.29**	**0.64**
MRE	0.55 (0.30-0.80)	0.84	0.43 (0.13-0.73)	0.82	0.44 (0.17-0.71)	0.78	**0.47**	**0.81**
LRF	0.55 (0.29-0.80)	0.84	0.51 (0.20-0.82)	0.86	0.42 (0.14-0.70)	0.78	**0.49**	**0.83**
MSF	0.66 (0.37-0.95)	0.84	0.60 (0.25-0.95)	0.84	0.66 (0.37-0.95)	0.84	**0.64**	**0.84**
MSE	0.21 (0.01-0.41)	0.52	0.14 (-0.18-0.46)	0.56	0.11 (-0.25-0.47)	0.56	**0.15**	**0.55**
MSR	0.50 (0.29-0.71)	0.74	0.43 (0.18-0.68)	0.72	0.41 (0.18-0.64)	0.70	**0.45**	**0.72**
SLS	0.27 (-0.15-0.69)	0.72	0.36 (-0.02-0.74)	0.72	0.64 (0.32-0.96)	0.84	**0.42**	**0.76**
ODS	0.69 (0.35-1.00)	0.88	0.72 (0.46-0.98	0.88	0.53 (0.34-0.71)	0.80	**0.64**	**0.85**
Mean	**0.54**	**0.79**	**0.47**	**0.77**	**0.44**	**0.74**	**0.48**	**0.77**
SD	**0.20**	**0.12**	**0.22**	**0.13**	**0.23**	**0.13**	**0.20**	**0.12**

CI= confidence interval; C Flex= cervical flexion; C Ext= cervical extension; C Rot= cervical rotation; MRE= medial rotation extension; LRF= lateral rotation flexion; MSF= multisegmental flexion; MSE= multisegmental extension; MSR= multisegmental rotation; SLS= single leg stance; ODS= overhead deep squat; SD= standard deviation

DISCUSSION

The purpose of this study was to examine the reliability of scoring the 10 top-tier SFMA tests using real time scoring in a group of healthy participants. The goal was to blend methodology from the two previous SFMA reliability studies[1,2] in order to examine the repeatability of the assessment. Overall, the results showed slight to substantial reliability and this compared similarly to previously published SFMA reliability studies.[1,2] Because the previous SFMA reliability study on a healthy population used video analysis, this study aimed to reproduce these

Figure 4. *Cervical rotation test.*

methods but score all participants real-time to simulate how the assessment is typically used in clinical practice. Conversely, the only other SFMA reliability study[1] examined a clinical population with a known pathology. While this replicates the intended purpose of the assessment, pain and/or dysfunction has the potential to affect the reliability. In theory, movements are more likely to be repeatable in an otherwise healthy population because there is less likelihood of pain affecting the movement pattern.

Both the FMS™ and the SFMA are movement systems that allow the clinician to evaluate movement quality. Cook[3] suggests that clinicians are knowledgeable in both techniques to understand how the two systems can complement one another. The SFMA is designed to be a systematic method for observing movement pattern dysfunction in a pathologic population to determine the root cause of the pain and/or dysfunction. These results can then guide potential clinical interventions. Conversely, the FMS™ is designed to be utilized with healthy participants in an attempt to guide programming decisions and provide movement feedback. Despite these differences, the two systems still share many similarities in their approach to assessing movement and two of the movement patterns are nearly identical.

A search for articles involving the SFMA revealed less than five peer-reviewed journal articles. As previously mentioned, several differences in methodology existed between the current study and the two published articles. First, both previously published studies used video recording for scoring participants. All three raters used video recordings to score participants in the Glaws et al.[1] study, while Dolbeer et al.[2] had two raters score real time and one score using video. While this is feasible for conducting research, Dolbeer et al.[2] acknowledged that the SFMA is intended to be scored real time in the clinical setting and not intended for 2-dimensional video analysis. Second, Dolbeer et al.[2] used a clinical physical therapy population, therefore, some of their participants exhibited painful patterns. Both Glaws et al.[1] and the current study used a healthy population and no painful patterns were recorded. Third, both Glaws et al.[1] and the current study examined both intra- and inter-rater reliability while Dolbeer et al.[2] only evaluated inter-rater reliability. In many clinical practices, the same clinician works with the patient and tracks movement progress, making it necessary to establish the intra-rater reliability. However, having the ability to allow another clinician to evaluate movement performance allows the system to become more versatile within a clinical practice with multiple clinicians sharing the patient load. Fourth, both the current study and Glaws et al.[1] used examiners with differing levels of experience with the SFMA while Dolbeer et al.[2] used examiners with similar levels of experience. While data from the current study may have been different if all of the examiners had the same level of experience, it is unlikely clinicians working together would all have the same level of experience. Finally, both previously published studies[1,2] used both categorical and the criterion checklist scoring tool that is included in the SFMA manual[4] to assess reliability. The current study only evaluated the reliability of the categorical scoring method. Additionally, the authors chose to only examine the categorical scoring method since these results determine if it is necessary to perform the breakouts associated with each of the movement patterns.

Despite these differences, the results demonstrated slightly lower intra-rater reliability and slightly higher inter-rater reliability when compared with Glaws et al.[1] Furthermore, the results demonstrated similar to slightly higher inter-rater reliability depending on which set of examiners were being compared. The current study found similar to slightly higher inter-rater reliability when comparing the Dolbeer et al.[2] study. However, some of this variation could be attributed to the presence of dysfunctional and/or painful movement patterns in the patients' evaluated in both studies. Based on how the Cohen's Kappa statistic is calculated, the Kappa coefficient can be lowered with small samples of a given categorical score.[10] Because the distribution of FN, DN, FP, and DP scores for some movement patterns were not evenly distributed for various movement patterns, the overall interpretation of the Kappa statistic may have been lowered. The purpose of the Kappa statistic is to take into account the possibility that raters guessed on scores.[8] With unequal distributions of scores, the Kappa statistic may have the potential to be excessively lowered.[8] This may explain why there were large differences in the Kappa statistic and the absolute agreement for some of the movement patterns.

Intra-Rater Reliability

The intra-rater reliability Kappa coefficients for all raters in the Glaws et al.[1] study ranged from 0.25 to 0.94 for the 10 fundamental movements. The most reliable rater in the Glaws et al.[1] study had Kappa coefficient values ranging from 0.41 to 0.94. The results from the current study demonstrated intra-rater reliability Kappa coefficients for all raters ranging from 0.21 to 1.00. The most reliable rater from the current study demonstrated Kappa coefficients ranging from 0.38 to 0.90. Values for absolute agreement from the Glaws et al.[1] study ranged from 0.63 to 0.97 while the results from the current study ranged from 0.64 to 1.00. While these ranges are strikingly similar, examining the individual movements revealed some substantial differences. For example, by averaging the results for all raters, cervical extension demonstrated the lowest intra-rater reliability (0.30) in the current study while this was the most reliability movement (0.77) in the Glaws et al.[1] study. Alternatively, multisegmental extension was among the least reliable movements in both studies with Kappa coefficients ranging from 0.36 to 0.49. Similarly, cervical rotation demonstrated lower reliability values while multisegemental flexion revealed higher reliability values. While the overall reliability results were similar when averaging all movement patterns, the two studies found discrepancies in the reliability for certain movement patterns.

Inter-Rater Reliability

Results from the current study showed inter-rater reliability Kappa coefficient values ranging from 0.11 to 0.89 for all raters. Glaws et al. reported values ranging from 0.07 to 0.97 for all raters while Dolbeer et al. reported values ranging from 0.1 to 0.96. Raters A and B displayed the highest reliability while raters B & C were the lowest. These data were similar to the Glaws et al.[1] study which also found level of experience was reflective of the inter-rater reliability. For all sets of raters in the current study, multisegmental extension exhibited the lowest inter-rater reliability, while cervical flexion demonstrated the highest. These same results were not found in the two previously published reliability studies.[1,2] Inter-rater reliability between sets of examiners revealed inconsistent results for the highest and lowest reliability values.

Top-Tier SFMA Movements

Regardless of the movement pattern being evaluated, clinicians must evaluate the pattern for excessive effort and/or lack of motor control. While this tends to be a subjective component of the evaluation, with practice, clinicians are better able to pick up on the cues associated with both excessive effort and lack of motor control. The current results demonstrated the lowest reliability numbers for cervical extension, cervical rotation, and MSE. Of these movements, the cervical rotation movement contains the most objective criteria by providing the midpoint of the clavicle as the target anatomical landmark. To be considered FN, the participant must rotate only the cervical spine so that the chin is aligned with the mid-point of the clavicle. While this movement contains objective criteria, it does require the examiner to reliably assess the mid-point of the clavicle. In contrast, both the cervical and multisegemental extension patterns require the examiner to assess the participant for a 'uniform' spinal curve during the movement. While this is only one component of the evaluation criteria, the subjective nature of assessing a uniform spinal curve likely contributed to lower reliability scores for these movements.

The movements with the highest reliability tended to have the most objective scoring criteria. For example, in order to be considered FN for the cervical flexion pattern, the participant must flex the neck and touch the chin to the chest. Similarly, the multisegmental flexion test contains several objective criteria such as a touching the fingers to the ground, a posterior weight shift, and a sacral angle of greater than 70°. Interestingly, the multisegmental flexion test also requires the participant to have a uniform spine curve, however, the reliability of assessing this motion was considerably higher than the multisegmental extension movement. A possible explanation for this could be the difficulty in visualizing the lumbar spine extension due to the reduction of this space during extension. Furthermore, not being able to touch the toes during the multisegmental flexion test immediately places it in the dysfunctional category. It is possible the reliability of the multisegmental flexion test is higher because the rater knowingly or unknowingly places more emphasis on evaluating the objective portions of the test and less on the subjective portions.

Limitations

As with all research investigations, this study was not without limitations. All participants were healthy at the time of testing; therefore, no painful patterns were identified. The SFMA is designed to identify dysfunctional movement patterns in participants with known musculoskeletal injury.[3] Secondly, the authors chose to assess only the intra- and inter-rater reliability of the categorical scoring system of the SFMA. The system allows clinicians to score participants based on both a categorical and criterion checklist system. The authors also chose to evaluate only the reliability of the categorical system because this system dictates what movements receive further attention. Third, movements were scored real-time by all participants. This reflects the intended design of the assessment; however, it does not allow evaluators to watch the movement multiple times. However, all evaluators had the ability to ask the participant to perform the movement again, which allowed the evaluator to take a second look or evaluate the movement from a different angle. It is possible the individual only narrowly passed/failed the movement during the first session or attempt and the opposite interpretation was found during the second visit or attempt. Naturally, this would negatively affect the reliability but does reflect the same challenges a clinician would find in clinical practice. Fourth, the current study utilized a sample of healthy, college-aged participants which may make it difficult to generalize the findings to other populations. Finally, examiners differed in their experience level with the assessment. While it may be helpful to know the reliability of a homogenous set of examiners, many clinical settings where the SFMA is performed contain a variety of clinicians with varying levels of SFMA experience.

CONCLUSIONS

The top-tier movements of the SFMA scored categorically and assessed real-time in a healthy population demonstrated slight to substantial reliability. The methodology of this study combines aspects of previous SFMA reliability studies and further supports their findings. Both intra- and inter-rater reliability was highest for raters that had the most experience and were certified in the SFMA. It appears movements with the most objective scoring criteria produce the highest reliability values.

REFERENCES

1. Glaws K, Juneau C, Becker L, et al. Intra- and inter-rater reliability of the selective functional movement assessment (SFMA). *Int J Sports Phys Ther*. 2014;9(2):195-207.

2. Dolbeer J, Mason J, Morris J, et al. Inter-rater reliability of the selective funcional movement assessment (SFMA) by SFMA certified physical therapists with similar clinical and rating experience. *Int J Sports Phys Ther*. 2017;12(5):752-763.

3. Cook G. *Movement: Functional movement systems screening-assessment-corrective strategies.* Ramsey, NJ: On Target Publications; 2010:311.

4. SFMA. Selective functional movement assessment seminar manual. . 2011.

5. Wainner R, Whitman J, Cleland J, et al. Regional interdependence: A musculoskeletal examination

model whose time has come. *J Orthop Sports Phys Ther*. 2007;37(11):658-660.

6. Bonazza N, Smuin D, Onks C, et al. Reliability, validity, and injury predictive value of the functional movement screen. *Am J Sports Med*. 2016;45(3):725-732.
7. Cuchna J, Hoch M, Hoch J. The interrater and intrarater reliability of the functional movement screen: A systematic review with meta-analysis. *Phys Ther Sport*. 2016;19:57-65.
8. McHugh M. Interrater reliability: The kappa statistic. *Biochemia Medica*. 2012;22(3):276-282.
9. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.
10. Byrt T, Bishop J, Carlin J. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993;46(5):423-429.
11. Padua D, Boling M, Distefano L, et al. Reliability of the landing error scoring system-real time, a clinical assessment tool of jumplanding biomechanics. 2011;20(2): 145-156. *J Sport Rehabil*. 2011;20(2):145-156.
12. Ekegren C, Miller W, Celebrini R, et al. Reliability and validity of observational risk screening in evaluating dynamic knee valgus. *J Orthop Sports Phys Ther*. 2012;39(9):665-674.
13. Onate J, Dewey T, Kollock R, et al. Real-time intersession and interrater reliability of the functional movement screen. *J Strength Cond Res*. 2012;26(2):408-415.
14. Smith C, Chimera N, Wright N, et al. Interrater and intrarater reliability of the functional movement screen. *J Strength Cond Res*. 2013;27(4):982-987.
15. Teyhen D, Shaffer S, Lorenson C, et al. The functional movement screen: A reliability study. *J Orthop Sports Phys Ther*. 2012;42(6):530-540.
16. Guglin H, Hoogenboom B. The functional movement screen (FMS): An inter-rater reliability study between raters of varied experience. *Int J Sports Phys Ther*. 2014;9(1):14-20.
17. Gribble P, Brigle J, Pietrosimone B, et al. Intrarater reliability of the functional movement screen. *J Strength Cond Res*. 2013;27(4):978-981.
18. Schneiders A, Davidsson A, Hörman E, et al. Functional movement screen normative values in a young, active population. *Int J Sports Phys Ther*. 2011;6(2):75.
19. Shultz R, Anderson S, Matheson G, et al. Test-retest and interrater reliability of the functional movement screen. *J Athl Train*. 2013;48(3):331-336.
20. Parenteau-G E, Gaudreault N, Chambers S, et al. Functional movement screen test: A reliability screening test for young elite ice hockey players. *Phys Ther Sport*. 2014;15(3):169-175.
21. Lucas N, Macaskill P, Irwig L, et al. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63:854-861.

APPENDIX A

SFMA Instructions

The starting position for all tests is the same. You will stand erect with your feet together, toes pointing forward, and your arms hanging comfortably by your side. We will read you the instructions and then ask you to perform the movement so we can score the quality and quantity via visual inspection. If you experience pain anywhere in your body during the movement, please let the examiner know.

1. **Cervical Spine Flexion**: Please attempt to touch your chin to your breastbone (sternum) while keeping your trunk erect during the movement.

2. **Cervical Spine Extension**: Please extend your head back like you are looking at the ceiling as far as possible while keeping your trunk erect during the movement.

3. **Cervical Spine Rotation**: Please rotate only your head as far as possible to the right. Repeat this movement to the left.

4. **Upper Extremity Pattern 1**: Please reach back with your right arm and try to touch where I have placed my finger on your left inferior scapula. Repeat this movement with your left arm, touching the spot I have marked.

5. **Upper Extremity Pattern 2**: Please reach overhead with your right arm and try to touch where I have placed my finger on your left scapular spine.

6. **Multi-Segmental Flexion**: Please bend forward at the hips while reaching your hands toward your toes. Try to touch the tips of your fingers to the end of the toes without bending your knees.

7. **Multi-Segmental Extension**: Please raise your hands above your head with your arms fully extended and the elbows in-line with the ears. Bend backward as far as possible, making sure your hips go forward and arms go back simultaneously.

8. **Multi-Segmental Rotation**: Please rotate the entire body (hips, shoulders, and head), as far as possible to the right while the feet position remains unchanged. Repeat this movement but rotate to the left.

9. **Single Leg Stance**: Lift your right leg so the hip and knee are both at 90-degree angles. Remain in this posture for 10 seconds. Rest for as long as you need and then repeat this task with your eyes closed. Repeat both tests with lifting your left leg.

10. **Overhead Deep Squat**: Please stand with the instep of your feet in vertical alignment with the outside of the shoulders. Your feet should be pointing straight forward. Extend your hands overhead with the shoulders flexed and abducted and the elbows fully extended (this will be demonstrated). Slowly descend as deeply as possible into a squat position. Maintain the heels in contact with the floor, your head and chest should remain facing forward, and the arms maximally pressed overhead. Return to the starting position.